MANAGING PEAK DEMAND FOR PASSENGER RAIL: A LITERATURE REVIEW

32nd Australasian Transport Research Forum (ATRF), Auckland New Zealand, 2009
Chris Hale and Phil Charles[1], Centre for Transport Strategy
The University of Queensland, Brisbane, Australia

ABSTRACT

A review of the literature on the phenomenon of peak rail demand and its management.

Major cities in Australia are experiencing heavy demand growth for rail passenger transport, especially in the morning and afternoon peak periods during which radially-configured rail systems experience heavy commuter traffic catering to the standard workday.

Passenger rail systems customarily focus attention on meeting peak demand – through investment in infrastructure, rolling stock and provision of services. Allocating extensive resources for use during only a few hours each day is not efficient.

The paper looks at the literature on peak rail demand. Firstly, the phenomenon itself is described, along with the methods that might be used to track and measure demand levels. Secondly, common and potential peak spreading or shifting or demand management approaches, issues and ideas are reviewed – including: network planning responses, differential pricing, policy and strategy and passenger location-balancing strategies. The relationship between the standard working day and peak demand is also reviewed.

Keywords: transport planning, rail demand management, peak period, crowding, mass transit capacity

Acknowledgement: *This paper was produced under Project No R1.107 Urban Rail Demand Management Strategies funded by the CRC for Rail Innovation (established and supported under the Australian Government's Cooperative Research Centres program).*

INTRODUCTION

This paper reviews literature relevant to the “peak periods” for passenger rail travel, and seeks to draw out the range of issues and potential management and strategy-based responses that can assist in alleviating problems associated with excessive peak loadings. In order to establish the environment in which *peak period demand management* operates, the following topics are covered:

A) Characterising and describing the peak period
B) Key performance indicators
C) Configuration of rail networks
D) The nature of commuter travel
E) Pricing and ticketing
F) Summary of challenges associated with peak demand profiles
G) Recommendations for addressing peak congestion

Terminology and descriptive techniques for discussing and analysing the peak period are provided initially, along with discussion of the various elements contributing to “capacity constraints” in rail systems against which peak period demand levels are matched.

In part B a list of key performance indicators is provided. These are drawn from the literature, and even at the early stage of research that this literature review represents, these indicators are of key interest in analysing and understanding the characteristics of the peak period from both a capacity and a demand point of view.

In part C, rail network configuration is discussed. It may initially seem tempting to look at peak demand responses from an isolated perspective and insist that any strategy is about “working with the existing system”. But analysis of the literature seems to be suggesting that network configuration characteristics are a prime issue in travel demand outcomes. Two attributes are of interest – the “planning responsiveness” of a system toward meeting changing population and travel demands over time, as well as the actual “network configuration”, both current and planned.

---

1 Chris Hale is a researcher and Professor Phil Charles is Director of the Centre for Transport Strategy, School of Civil Engineering, The University of Queensland

In part D, we diverge into a discussion of the attributes and characteristics of work-related public transport travel. Being informed by real-world travel realities allows us to consider which responses to peak period congestion are likely to meet the readily observable preferences and needs of commuters and other travellers

Part E features discussion of a cornerstone travel demand management tool – *pricing*. In this section the "four attributes" of efficient pricing are identified – distance, time of travel, choice of station, and choice of line or service. The relationship between pricing and ticketing technologies is an emerging frontier and some key issues are discussed.

Part F allows us to pause before final recommendations. At this point we review some of the main reasons that excessive peak period congestion is problematic from strategic, cost and revenue points of view. High levels of utilisation of peak period services may overshadow the inherent inefficiency of overly peak-loaded systems, in which every fully-loaded train is offset by empty seats and inefficient utilisation of non-peak services.

The final section of the paper summarises key areas of opportunity for addressing peak period rail demand.

In terms of methodology, the literature review started with an attempt to reasonably comprehensively cover the existing literature in what is only an "emerging" field, rather than a well-established discipline with a clearly-defined body of literature. Emphasis was given to currency of publications reviewed (papers with publication dates in the previous 10 years were prioritised). Overall, because of the limited existing body of literature, the number of publications reviewed was relatively constrained – but the analytical quality of many of the sources seems to have compensated effectively for any missing breadth. Additional sources of information were brought in from topics or fields that were considered "adjacent", or directly relevant (for example – publications on ticketing and pricing, and the commuter experience, plus a handful of publications recognised as leading sources in contemporary mass transit planning).

A. CHARACTERISING AND DESCRIBING THE PEAK PERIOD

"The basics of rail transit capacity are very simple – the product of how many trains can be operated in the peak hour and by the number of passengers that will fit on those trains." (TCRP13 1996, p xi)

Part of the initial process toward tackling peak period rail congestion problems may involve formulating a more effective descriptive toolkit to outline and discuss the issues being encountered.

Terminology

Some rail agencies describe the morning peak period as running for up to 3 or 4 hours (say 6am to 10am) – and in a general sense this is probably true, in that demand levels during this extended period are very often well above passenger demand levels outside this period. Whether this suffices for an effective description and for addressing peak-related problems is another question entirely. At the most simplistic level we need to ask ourselves whether encouraging travel between, say 6am and 7am, is not a better outcome in terms of capacity constraints compared to catering to more travel between 7am and 9am. In most cases the earlier timeslot will either have available capacity, or less severe overcrowding.

Although US planners and sources have traditionally referred to the "peak 15 minutes" (TCRP13 1996) this could be seen as peculiar to the operating regimes and prevailing thinking in US transit systems. These conditions and regimes generally tend to feature low levels of frequency, lower mass transit patronage and as a consequence a desire to handle the bulk of peak period patronage in a single short burst or "wave" – or in some extreme cases with a single train. By contrast, Australian or other passenger rail systems may well see a "peak 15 minutes" and a passenger "wave" at some point, but the importance of this sub-period relative to other times throughout the morning or evening *peak hours* is much less prominent and relevant.

In the USA …*"Commuter rail scheduling is often tailored to the peak travel demand rather than operating a consistent service throughout the peak period."* (TCRP13 1996, p5)

More useful terminology refers to the "peak", "shoulder" and "off-peak" periods. This is not perfect but at least carries some level of nuance and practical value. Yet to appear in widespread descriptive use is the demand characteristic of a "peak within the peak" (TCRP13 1996). This would normally occur close to 9am in a CBD destination station catering to high levels of commuter patronage. The “peak within the peak”, referred to here as the “spike”, is a significant rail demand phenomenon firstly because it is this concentrated sub-period of the peak during which the most severe overcrowding is occurring – along with related consequences such as passengers being left behind and experiencing travel delay. Secondly, and perhaps even more significantly, the short and concentrated period of “spike” demand then becomes the first trigger-point for any requirement to upgrade capacity. As many capacity upgrades in CBD locations will inherently require expensive infrastructure, rolling-stock and service-provision responses, the concentrated “spike” period is likely to disproportionately impact on system costs. This is especially problematic when “step functions” are faced, in which locations already operating at the limit of available system capacity need to implement a paradigm increase in capacity, rather than incremental upgrades.

Workable descriptors

A sorting and summary of different definitions (ATC 2005; Li & Wong 1994; Doggett et al 2004; TCRP13 1996) based on their practicality and informative power for these periods might suggest the following break-down of timings and periods could be workable:

Prior to 6am:	*off-peak morning*
6am to 7am:	*morning shoulder*
7am to 8am:	*morning peak*
8am to 9.15am:	***morning spike***
9.15am to 10am:	*late morning shoulder*
10am to 4pm:	*midday off-peak*
4pm to 5pm:	*afternoon peak*
5pm to 6.15pm:	***afternoon spike***
6.15pm to 7.30pm:	*evening peak*
7.30pm onward:	*off-peak evening*

These descriptors will be referred to subsequently throughout this paper.

Station capacity

> *“The most common constraint is the close-in movement at the maximum load point station.” (TCRP13 1996, pxii)*

Discussing the different aspects of rail transit capacity may clarify a range of rail demand-related issues. High on the list of priorities for most rail systems is the level of demand being encountered at individual stations in the network (Vuchic 2005, 2007). The key question is whether a particular station is designed and configured in a manner that allows for safe, convenient and reliable handling of the large volumes of passengers using that facility – particularly during peak periods. High passenger volumes will need to be handled in a manner that attains accepted benchmarks and standards for quality of service (TCRP100 2003).

Transport experts have described station capacity as among the most significant constraints that a system faces in handling ever-higher patronage levels (TCRP13 1996). Because they are fixed infrastructures with pre-set configurations and restricted real estate footprints, there is almost no ability to expand station capacity to meet growing peak demand levels in a short-term operational sense. In the medium term, problematic levels of demand resulting in overcrowding may be met with a range of station capacity-enhancing measures, either focused on throughput of trains, or more effective circulation and ticket-processing measures for passengers. In some scenarios wholesale expansion of a station facility may be required.

> *“If platforms are too narrow, or exit paths limited, congestion on the platform can cause delays in unloading a train; this can affect the overall station dwell.” (TCRP13 1996, p12)*

Alternatively, non-infrastructure demand management tools and techniques applied either on a system-wide basis, or directed at a particular line or station could offer some hope of managing excessive levels of demand at key stations. Demand management techniques might be brought into play whether station capacity and demand is immediately problematic or not, but their need and importance in identified situations of overcrowding and excessive peak patronage seems clear.

Train capacity

Train capacity is another salient factor in the discussion of overcrowding and peak period demand. Clearly, any surge in demand during particular times of the day has the potential to overwhelm the carrying capacity of scheduled services – with passenger loadings on individual trains an issue of particular interest. Train capacity is comprised primarily of the number of cars in a train set, their seating capacities and standing capacities (TCRP100 2003; Vuchic 2007). This "total" train capacity figure is a key component within the overall contributors to system capacity. Potential capacity-enhancing measures theoretically include using longer trains, double-decked trains, and altering the ratio of seats versus standing capacity in favour of increased standing volumes (lifting overall train capacities). These potential measures are "theoretical" because in the short-term there is limited opportunity to make any changes. The changes are also costly, even where sufficient time is available to make them. Increasing train capacities may be counterbalanced by resultant passenger loading-time problems, or may result in unpopular increases in standing time for longer journeys. And finally, in many systems carrying high volumes, most of the readily available measures for boosting train capacities may already have been implemented.

Any demand management measure offering the ability to handle high passenger volumes without altering or acquiring rolling stock or infrastructure at significant cost is likely to be an efficient and effective option.

Line capacity

"Line" capacity is the final significant element of overall system capacity, which may come under pressure from high levels of peak period demand (Vuchic 2005). Line capacity represents the number of trains that can be moved through a rail corridor (or single line) over a set period (normally measured by the hour). In this sense, there are a range of "above track" (operational and service-based) and "below track" (infrastructure and fixed systems) aspects that contribute to overall line capacity. These would include the quality of alignments and allowable speeds determined by track quality (Vuchic 2007). It would also include signalling and train control systems – with potentially 20% or greater line capacity improvements available through the use of "moving block" train control systems (TCRP13 1996). These systems are not in widespread use in Australian networks.

There is also the question of optimising the interactivity between track and signalling systems and rolling stock – with the basic question of "where does the performance constraint lie". Modern rail networks (particularly in Europe and Japan) are quite capable of moving trains at high levels of speed and performance when the interface between rolling stock and fixed network elements are inter-operating and optimising effectively. This implies that advanced systems offer greater carrying capacity – so a pertinent question for railways catering to heavy levels of demand is the extent to which they are optimising the interactions and performance of different system elements in order to contribute to a stronger overall line capacity (Vuchic 2005).

All of the above and below track measures for boosting line capacity involve considerable expense and lead-times.

System capacity

"Rail transit capacity is set by the weakest link or bottle-neck on a system." (TCRP13 1996, pxx)

System capacity is the overall peak carrying capacity of corridors and entire networks, based on the effectiveness or otherwise of the different contributing elements – station, train and line capacities (Vuchic 2005). In most cases, concerted effort at optimising these contributing elements could yield increased capacity in the medium and longer term. In certain systems however, particularly problematic components of the network requiring large infrastructure-based responses will be a brake on capacity expansion. The "weakest link" generally sets the limits of system capacity. Peak demand management approaches are of interest particularly to the degree that they can lift performance, relieve pressure and/or obviate problems associated with the "weakest links" in a particular network, as a first stage before expensive and long-term infrastructure based solutions.

Keeping the problem within manageable limits

European transport planning sources and methods provide insights and ideas, and offer promise for a future in which rail demand is more actively tracked and managed. From the rail planners of Munich, a city with a deserved reputation for excellence in rail infrastructure and management, a managerial tool of a "mandated capacity ceiling" (in English translation) presents itself - and is evidently in use in Munich's U-Bahn subway system (City of Munich 2005). This tool tracks demand against capacity through the day and sets a *mandated* ceiling – a limit or level at which rail patronage demand levels in excess of the ceiling automatically triggers a planning or demand management response from local rail planners. The *ceiling* becomes a trigger point at which patronage and vehicle occupancies beyond a certain pre-set level are seen as *requiring* a management response.

> *"Although rail transit is noted for reliable and regular operation, minor delays are routine and an operating margin (is)…essential to prevent delays from compounding." (TCRP13 1996, pxi)*

In Munich, the maximum level for the "mandated ceiling" is set at 65% patronage against total offered capacity during all periods, including the peak. This seems to be a low mandated *ceiling* for capacity/utilisation, however Munich's transit system is performing very well in economic terms (so the benchmark is unlikely to be a "mistake"). Munich is also a large and diverse transit system which allows excessive crowding to be reasonably easily catered to through enticing passengers into other service options. On these counts, the approach adopted in Munich is worth observing and understanding, but is also possibly beyond the practical and immediate reach of Australian systems in which particular rail corridors are often the only viable transit option available to commuters. While the application of "mandated ceilings" should be explored for application in Australian cities, it is unlikely that a 65% loading figure is workable or affordable.

Vuchic (2005) and other suggest individual lines should not exceed 80% passenger loadings, due to the inevitable reliability and comfort-related problems that these high-end loadings involve. A second reason commonly offered for the need to buffer train capacity is the "*loading diversity factor*" (TCRP13 1996; TCRP100 2003) – which recognises the reality that loading trains at an "optimal" or full-capacity level is unlikely, due to the semi-random and unequal distribution of passengers with carriages throughout a train.

Overall, greater attention to *demand management and tracking tools and techniques* seems to be presenting itself as an important option.

B. KEY PERFORMANCE INDICATORS

In addressing peak period management scenarios and options, a set of basic performance indicators can provide an effective "snapshot" of the system and its ability to cope with peak demand levels, as well as being able to ascertain the overall "peakiness" of a particular system. Drawn from a range of sources (Vuchic 2005, 2007; TCRP88 2003), the following represents a list of important indicators that can offer an understanding of peak-period issues:

Passenger movement capacity/utilisation indicators

- Train capacities and loadings – seated, standing and total
- Line capacities & volumes – in terms of; available paths versus utilised paths
- System carrying capacities & ridership - theoretical, offered, and utilised - based on individual lines and the entire network. Also expressed as a ratio (volume/capacity)
- Average trip length per passenger on the system and on key lines
- Standing time duration, and ratios of passenger km to seat km
- "Pass-ups", or the regularity at which passengers are unable to board due to overcrowding

Station indicators

- Numbers boarding, alighting and transferring at particular stations – daily, and either in one hour blocks or 15 minute interstices within peak periods
- Station passenger-handling capacities

Ticketing, fares, cost and revenue indicators

- Peak fare versus off-peak (and whether in place)
- Relative usage levels of various fare options
- Revenue per run or km (and profitability) on key routes during peak and off-peak, in both directions
- Cost per passenger (system average)
- Farebox recovery ratio

Demand levels over time

- Accurately measuring and analysing peak v off-peak travel, as well as variations due to weekly and seasonal factors
- "Peak to base ratio" – mapping the differences between peak loadings and averages throughout the day and reflecting these important and complex issues through a single figure or "score" with strong descriptive impacts. This is also known as the "*coefficient of flow variations*" (Vuchic 2007)
- Percent of "reverse-commute" trips

Policies and strategies

- Customer attitudes and issues (by survey) related to peak period travel experiences, as well as reasons for peak travel, and propensity to use off-peak periods depending on different service and fare options
- Description of peak period management approaches in place. Qualitative analysis of outcomes where possible.

C. CONFIGURATION OF RAIL NETWORKS

"Observing general transportation trends in different countries, it can be easily seen that the growth of cities increases volume and density of travel in urbanized areas. How much of that increase goes to transit modes depends on auto ownership level, ***transit extensiveness, service quality, and efficiency****, which are largely functions of transportation policies in different cities and countries." (Vuchic 2007, p432)*

A variety of sources (Brooker & Moore 2008; Cervero 1998; Vuchic 2005, 2007; Doggett et al 2004) have suggested that the overall configuration or layout of a particular rail network and its relationship to activity and population centres and travel generators is of particular interest in understanding problems of congestion.

Planning responsiveness

Rail network congestion is generally evident on particular lines and at particular stations. A significant question in addressing rail congestion outcomes and developments, is whether an existing network is able to cater to reasonably foreseeable levels of travel demand in particular sectors and to and from important locations.

"…lack of compatibility between passenger needs and management's perception of those needs could result in the misallocation of scare resources as well as growing passenger dissatisfaction with transit services." (Thevathasan & Balachandran 2007, p1)

At the heart of the issue of peak period management in Australia at present is the significant and ongoing growth in rail passenger travel, particularly for work-related journeys, that has been occurring in recent years (Brooker & Moore 2008). Matching of metropolitan travel needs with infrastructure and service provision over time might be characterised as the *planning responsiveness* of a rail network. In this aspect we are interested in whether the planning and development of a rail system has kept pace with the population growth and travel demand characteristics of the city it serves. Mismatch between population and economic growth, travel demand and travel desire lines, and "planning responsiveness" of a system over time, is likely to result in excessively peak-loaded demand levels and unacceptable levels of rail congestion.

"Since 2006, increased rates of growth of peak hour rail passenger travel, driven by increasing road congestion, greater environmental awareness and most recently, spiralling fuel process, have exacerbated existing peak hour rail network capacity limitations."

(Brooker & Moore 2008, p203)

Network configuration

"The Sydney passenger rail network is heavily focused on the Sydney CBD."

(Doggett et al 2004)

Network design or configuration is the degree to which the system layout provides effective *area coverage* (Vuchic 2005; Doggett et al 2004). Many networks have evolved into providing a unifocal or CBD-focussed network. This may arise because of legacy or historical network configurations, or where the system has not expanded into a true *network* configuration.

> *"With passenger volumes rather evenly spaced along the line and loads well distributed through most daily hours, average load factors on these lines tend to be high. This results in lower cost per passenger than on commuter-dominated radial lines. (Vuchic 2005, p205)*

Radial rail networks catering primarily to work commuting are more likely to produce problems of rail congestion and peak-loaded travel demand. This is less likely for *network* style systems that cater to a wider variety of travel opportunities, destinations, travel times and travel needs (Vuchic 2005).

> *"…peak-to-base ratio is the highest for radial travel dominated by commuting... (Vuchic 2005, p37)*

Rail development, planning and the urban growth histories of Australian cities have produced mass transit networks that are highly radial or unifocal, and planned and configured to cater mainly for work-based trips to the central business district. Both of these network attributes are considered to be prime *causes and contributors* to any problems of peak period congestion that are being experienced. Problems of unifocal systems are exacerbated if and when the distribution aspects of a mass transit network in CBD and central urban locations are also under-developed. Sydney and Melbourne, for example, at least spread peak demand across multiple stations within the CBD, while Brisbane has only the one true CBD station available (Doggett et al 2004; Thevathasan & Balachandran 2007). The peak smoothing problem then becomes one of better utilisation of existing infrastructure and stations first and foremost for these "advantaged" cities with more than one CBD station option (Doggett et al 2004).

Another longer-term question mark probably remains over the quality of infrastructure and service delivered to inner-city (but non-CBD) locations within Melbourne, Sydney and elsewhere (Brooker & Moore 2008; Thevathasan & Balachandran 2007). While Melbourne, as one example, has a highly important frame of inner suburbs, these economically and culturally important locations are not 'networked', but require radial in-out trips to the CBD in order to complete a particular journey. In this sense, the lack of network-style coverage of CBD-fringe and inner suburban locations may be exacerbating problem radial travel patterns, and denying development of these locations as employment centres that could conceivably reduce the strain on Melbourne's CBD-oriented peak-overloaded rail system. In this simple example, we see some of the problems associated with the overly-radial systems that are common in Australia (Brooker & Moore 2008; Thevathasan & Balachandran 2007).

Effective systems

> *"The ability of the existing heavy rail network to meet increased peak hour travel demand and still provide acceptable travelling conditions is now compromised in many locations." (Brooker & Moore 2008, p203)*

Contrasts can be drawn between Australian radial, CBD-focused, commuter-oriented rail systems versus the approaches, systems and traveller profiles of leading European cities. Both as a result of historic legacy and due to concentrated effort and investment, many European cities (London, Paris, Munich and Vienna to name but a few) have been able to deliver multi-destination, multi-station network-style systems that cater to a range of travel demands (Cervero 1998). These systems effectively spread the load across any number of major or minor rail stations in centrally located areas. While these cities are not without rail congestion problems, both existing networks, and the planning responsiveness and overall network development approaches are less inherently problematic and more diverse than those offered by the more unifocal or radial Australian systems. The Australian systems seem to be much more commuter-dependant and restricted in terms of the *number* of centrally-located rail lines and stations (Vuchic 2005).

Where to from here?

It seems that both the literature and the practice as observed in "successful" European rail systems suggests that operational and short or medium-term issues aside, Australian cities may need to fundamentally rethink the shape of their future rail and mass transit systems. Australian cities may also need to rethink the role of rail in serving multiple passenger market opportunities beyond the classic peak-period white collar CBD commuter. They may need to initially look at better network circulation elements and destination spreading through more and better-planned CBD stations, or through better utilisation of existing but less-popular CBD stations.

> *"…some distribution of patronage to lesser utilised stations particularly in the PM …may mitigate current and projected station capacity constraints in future." (Doggett et al 2004)*

Polycentric land use development, and transit oriented development in non-CBD locations also appear to be important medium to longer-term initiatives in order to spread demand levels and deliver smoother overall patronage outcomes in station and line loading, travel-timing, and direction-of-travel terms (Brooker & Moore 2008; Cervero 1998).

CBD-adjacent inner city locations are prime candidates for congestion-relieving network development and extension. Early-stage initiatives and network planning discussion in Sydney, Melbourne and Brisbane have all reflected some level of this thinking in recent years (Brooker & Moore 2008, p203). Sydney with multiple metro proposals, Melbourne with plans for CBD rail tunnelling and new stations and Brisbane's CBD metro rail proposal and associated new stations are all reflecting the idea of spreading the destinations while still responding to peak demand with an infrastructure response. The degree of success will likely be dictated by the quality of network design. Achieving a closer marriage between new infrastructure and service offerings on the one hand, and a broader set of demand-smoothing initiatives (non-infrastructure responses in particular) on the other, also presents itself as a key success factor moving forward.

From a cultural perspective, Australian cities are slowly beginning to shed the idea that "the rail system we have is the one that was provided 60 years ago". System expansion and development is back on the agenda. Success in rail planning and the livability of these cities into the future is likely to be a function of the extent to which planning is able to adopt successful European thinking that has seen the development of multi-destination networks serving a diverse array of passengers at all hours of the day and throughout the week.

Increased inner-city population and employment densities and a high propensity to use public transport among residents living close to the centre of most major Australian cities seem to be offering potential for resource-efficient rail network expansion.

In summary, "planning responsiveness" and "multi-destination networks" are longer-term tools for addressing problems of rail congestion during peak periods. The effectiveness of infrastructure-based approaches can be further enhanced with management-based approaches.

D. THE NATURE OF COMMUTER TRAVEL

In addressing peak period congestion issues, planners should be familiar with the characteristics of work-related rail travel, an influential factor both in terms of outcomes and problems associated with peak period congestion, and in the range and effectiveness of potential responses. In most cases, there is a "two-step" process required in order to change travel behaviour. The first "step" is recognising and discussing factors that genuinely limit people's travel time preferences to peak periods. The second "step" may lie in identifying which groupings of passengers and potential passengers actually have some scope to make a change, and then identify the circumstances or support mechanisms that are likely to encourage change.

> *"…income, habit and journey time are the most important variables determining the route choice." (Li & Wong 1994, p307)*

People sleep

In transport-realted contexts "normal" (or predominant) human sleep patterns involve around 8 hours of allocated sleep time roughly between the hours of 10pm and 7am. Some people may alter that allocation and timing to a degree, earlier or later, either willingly and freely, or out of some job-related or other need, but for the most part, people seem to prefer this timing due to its compatibility with ingrained social norms and human biology. There may be limited scope to shift commuter travel behaviour on a significant scale in a manner that would involve large numbers of people altering their preferred sleep patterns (Lyons & Chatterjee 2008).

Discussion and attention is sometimes turned to “shifting” travel choice into “shoulder” periods outside the morning (around 7-9am) and evening (4-6pm) peaks. As most people require around an hour or so of morning preparation after waking and will need to access their railway station and wait for their preferred train (perhaps a half-hour at least for these activities), catching a train *prior* to commonly-designated peak periods (usually starting at 7am) would likely involve pre-5.30am waking on a regular basis. It seems on the face of it somewhat unlikely that this would be of interest to large numbers of people or rail commuters. Research undertaken for CityRail during 2008 seemed to confirm this constraint to some degree (TNS 2008).

In summary, while a wide range of rail travel periods are viable and appealing, there may be a *restriction on the expectation of off-peak or shoulder-period travel ever becoming common or popular prior to, say, 7am*. Efforts aimed at shifting travel time into shoulder periods probably need to take account for this attribute of human behaviour.

People have jobs (and schools to go to)

“Work and school trips are less discretionary in time than other trips.” (Vuchic 2005)

There are also other socially oriented limitations on shifting commuter travel behaviour that need to be taken into account (Lyons & Chatterjee 2008). Standard office hours (9-5) are a reality of working life and seem to be entrenched for a number of reasons, including but not limited to: providing compatible and predictable hours for business-to-business relationships and customer service; allowing sufficient time in the morning and late afternoon for travel to and from work during daylight hours; and allowing predictable free-time periods in order to undertake scheduled leisure (TNS 2008).

“Statistics confirm that for many commuters there are other journeys that also need to be organized as part of the working day.” (Lyons & Chatterjee 2008)

Workers in certain industries and settings (particularly in the public service) already have access to flexible working hours. Research by TNS (2008) for CityRail in Sydney identified that *formal acceptance* of flexible working hours is not necessarily translating into genuine flexibility in many workplaces. Other workforce participants beyond the “white collar office worker” also use standardised hours. Many of these are already travelling outside the “peak” period for rail travel and should already be contributing to the smoothing of demand levels during the day. Tradespeople, as one example, are noted for their early starts and finishes relative to the office worker. These workers may be adversely affected where service levels in non-peak periods are not amenable to their travel needs – and use of private vehicles is an alternative in timings where road congestion is also less onerous. Students, both in high school and post-secondary education, are already living their lives in a timing arrangement that is staggered to the 9-5 office worker’s day. In large part they should theoretically already be contributing to smoother daily demand profiles (Li & Wong 1994, p314), but the extent to which students may actually be travelling during peak periods could be worthy of attention, due to their apparent propensity for non-peak travel as well as their apparent receptiveness to classic demand management tools such as differential pricing.

In summary, the degree that working (and school) hours and arrangements can contribute to smoother rail demand profiles across a given weekday seems to rest on (Lyons & Chatterjee 2008; TNS 2008):

- Formal programs for flexible work hours among major employers
- Acceptance and facilitation of those programs by participating employers
- Continued support by rail operators to encourage workers, students and others with “non 9-5” hours to use rail transit in non-peak periods
- Targeting off-peak travellers who are currently using cars, assuming the non-peak periods offer spare rail capacity
- Rail service levels in non-peak periods (Cervero 1990)

“Greater flexibility in working schedules might alleviate some commuting stress.” (Lyons & Chatterjee 2008, p186)

People have social and family routines and daily travel habits

Discussion of social constraints and family issues as they relate to preferred travel times is also warranted. Beyond fixed working hours, these issues seem to be major contributors to both the "limitations" people face in altering their timing of rail travel, as well as their potential motivations for doing so (TNS 2008; Lyons & Chatterjee 2008).

Recognising and supporting the opportunities, needs and desires of workers to allocate daylight and early evening hours to social, leisure and family activities seems to be a potential motivator for changing travel habits. Assuming workplace flexibility, many passengers may be attracted to the idea of travelling by rail-off peak if this is pitched as part of a desirable "package" or semi-regular "habit" of working non-conventional hours on particular days to leverage time for family, social activities, shopping or personal business. Service levels in off peak periods are a key determinant in supporting changed travel habits (Thevathasan & Balachandran 2007).

> *"…in comparison with time and habit, priced or fare differential is of less importance in determining the trip maker's route choice." (Li & Wong 1994, p314)*

Research on travel behaviour has pointed out that daily travel is for many people undertaken as part of a "routine" or "habit-based process" (TNS 2008). Workers will not necessarily want to choose a "new" travel route and timing each and every day – not least because of the effort required to check route maps, ticketing information and timetables. The potentially complexity of the daily travel scenario is simplified by most people into a routine that is easily remembered and adhered to.

> *"…many individuals may suggest that …choices are not available to them – they are bound into an habitual or routine commute pattern." (Lyons & Chatterjee 2008, p196)*

In short, the "habitual daily travel pattern" is an informed guess aimed at securing a higher likelihood of convenient travel (Doggett el al 2004; Li & Wong 1996). The salient point in this noted attribute of travel behaviour is that any attempt by rail agencies to shift daily travel behaviour will need to recognise that in most cases it is an ingrained and habitual pattern of travel that the agency seeks to change. Shifting ingrained behaviour will logically require support, reliability, an offer of additional convenience, and no small amount of marketing and information-provision from the rail agency in order to attract a traveller's attention to the idea of changing. All of these factors are then called into play again in order to effectively *cement* increased non-peak travel as a new habit (TNS 2008).

In meeting these needs, agencies should pay attention to the opportunity of boosting service levels in non-peak periods to better meet the demand for flexible travel options (TNS 2008), as customer responsiveness to improved off-peak service is significant (ATC 2006; Cervero 1990).

E. PRICING AND TICKETING

Pricing is seen as one the most effective options that rail agencies have in meeting the challenges of overcrowding and peak period congestion. Drawing from basic economic theory, the concept of rail pricing includes but seldom meets the premise that the full cost of service provision should be reflected in the cost to consumers (passengers). Like a range of other "public goods", rail passenger transport has seen its pricing mechanisms diluted and made subservient to a number of other socially and politically-oriented considerations (TCRP94, 2003).

All tickets are subsidised – but are they properly priced?

While a handful of international rail operators manage their business profitably or on a break-even basis due to effective management and robust ticket prices (Streeting & Charles 2006), many other agencies, particularly in the car-oriented "New World" have been running heavily subsidised and heavily politicised transit operations (TCRP94, 2003). This is in no small part due to the fact that road transport is also heavily subsidised, and as such any government support to mass transit is seen as a "second best option" in the absence of pricing car usage on a full-cost basis. Streeting & Charles (2006) identified a range of stated objectives from world rail agencies in their approach to setting fares – with revenue/financial position competing with many other issues for attention, and "managing peak period congestion" not raised as a substantial issue. Cervero (1990) also outlined a scenario in which cost/revenue outcomes were seen only as one concern competing with a variety of other issues and goals.

The policy environment in many cities has rendered rail heavily reliant on government subsidies for a number of decades now. This same period has seen a "stasis" emerging in the economic and management models used and in the overall outlook of mass transit as a business. Rail came to be seen as reliant on political patronage and oriented toward a "social support" role for "people who can't afford cars" (Vuchic 2005; TCRP94 2003). This variety of political, cultural, social and economic undercurrents plays no small part in the current difficulties for operators attempting to look at the pricing of their product in an objective manner, in conjunction with an understandable expectation from customers that only minor changes in price should occur at scheduled price-review intervals. But the strategic context in which the industry operates has moved on substantially and it appears increasingly likely that transit agencies may face a need in future to look at wholesale fare structure reconfigurations.

Rail mass transit is now popular among a broad range of target demographics, with the working commute market dominant (Brooker & Moore 2008). Rail is no longer for "people who can't afford cars" and the "average morning rail commuter" in many world cities is now notable for their high income and their willingness and ability to pay a higher level of marginal ticket cost in order to attain better levels of service and convenience (Cervero 1990). The widespread implementation of effective concession fare structures also means that re-addressing peak pricing in a rational manner is unlikely to involve the backlash and political fallout characteristic of US transit pricing debates in recent decades (TCRP94 2003). Cervero (1990) has already raised the possibility that ingrained assumptions on fares and pricing may be more influential than rational assessment or structured policy formulation approaches.

Any rational economic appraisal of the operating and strategic environments would lead to the conclusion that *pricing differentials are a cornerstone component of steps toward better management of peak period congestion*, and potentially of moves toward a more robust set of business conditions for passenger rail that sees the industry less reliant on government handouts with political strings attached (Cervero 1990). Pricing differentials are pricing *efficiencies* when done well. A more widespread and more sophisticated application of differentiated pricing should lead toward more efficient economic conditions for rail passenger transport – and alleviate peak period overcrowding, at least in part.

What exactly are we pricing?

"Fares differentiated by both distance …and time-of-day appeared to provide a balance of efficiency, equity and revenue benefits. (Streeting & Charles 2006, p3)

Another reason that inefficient, poorly-differentiated pricing structures have evolved for mass transit is the fact that pricing these services and pricing individual rides is inherently a complex task. Fare structure strategies need to first decide whether to account for *distance travelled* (Streeting & Charles 2006; Cervero 1990), and this is no small task. Even simple measures such as distance or zone-based fare structures add unwanted complexity to the decision-making needs of passengers and potential passengers. "Flat fares" have found favour in a number of locations (Streeting & Charles 2006; Vuchic 2005) because of their ease-of-use from both a passenger and an operator point of view – but this is not to say that ignoring distance in setting fares is not inefficient and inequitable. The encouragement of longer-distance travel at the same price as shorter distance rides is both illogical and widespread. The idea emerges that *any* pricing strategy is inherently complicated – let alone a strategy that broadens its remit of issues beyond distance and into the full range of cost and revenue implications in rail travel.

"…the common practice of flat fares is highly inequitable…." (Cervero 1990, p117)

Returning to the previous discussion of the nature of peak period travel may provide assistance and clarity – remembering that peak period "problems" are related to the carrying capacity of particular lines and corridors, the carrying capacity of certain stations, and the timing of travel for large portions of the commuter and other passenger markets. These three aspects of peak travel should provide clues as to the components of service that may benefit from an effective differential pricing regime. In summary, effective pricing may need to include differentials based on *distance* travelled, but also time of travel, choice of station and choice of line or service.

Li & Wong (1996) took a look at fare structures and ticket technologies – identifying the *potential* to price these attributes, as has Vuchic (2005). But few networks currently address *all* comprehensively. Elsewhere, analysis of "elasticities of demand" has offered the possibility that price-based initiatives can shift behaviour (ATC 2006), without going so far as to apply pricing in a structured manner to the 4 attributes of rail travel *cost*.

How does pricing relate to revenues and costs?

"Differentiating fares by peak and off-peak periods represents potentially the most effective way to capture the higher marginal cost of providing rush-hour services." (Cervero 1990, p128)

If rail agencies are willing to address pricing and fares in an objective manner, fares for some markets and locations may be "differentiated upward" in a manner that provides revenue growth opportunities. The rail operator faces the question of whether they are indeed willing and able to use pricing to generate stronger revenues, or whether they are simply seeking to re-price and redistribute travel while remaining under the same overall regime of subsidies and revenues. In some cases, including Hong Kong (Li & Wong 1996), US examples (TCRP94 2003), and the recent "smart saver" trial in Sydney, fare restructuring seems to be premised on either the *idea* of "revenue neutrality", or an apparent government mandate to remain revenue neutral (without elaboration on the justification for this approach).

"Overall, the potential efficiency and equity gains of differentiated pricing would likely far exceed any costs that might be associated with more complex fare collection or the loss of a few disgruntled customers." (Cervero 1990)

Agencies may also look at reducing fares for some travel choices as a means of attracting more patronage to that option (including by time, line and station) – which may assist in boosting off-peak travel and taking pressure off the most overloaded travel periods, routes and locations. In another application of the "second-best" principle, agencies could look to "differentiate downward", but only so far as that is logical and objectively supportable. Overall, transit agencies could consider using differential pricing to (Cervero 1990):

- Match-up on the different costs of catering to different travel choices
- Differentiate fares upward or downward based on these costs
- Seek to have pricing more closely related to cost, and to remove illogical cross-subsidy scenarios wherever possible
- Consider growing revenues through "upwardly differentiated" pricing (particularly in peak periods and locations).

Transit agencies then face the need to choose appropriate fare levels, which will relate to:

- Travel change generated by impact of a pricing change
- Willingness and ability to pay of customers
- Efficient use of any revenue growth resulting from changed fare structures.

The case made to customers and politicians for more logical and efficient fare structures that may be able to improve peak congestion on the basis of increased peak fares would rest on whether or not the additional revenue is put to good use. Some might argue that reducing absolute or percentage-based operating subsidies from government is a "good use" in its own right. Others, perhaps operating on the "second-best" principle related to the cross-elasticities and complicated subsidy regimes for car transport, might suggest that *improved* service, or *meeting peak-related infrastructure and service demands*, could be effective uses of new revenue from efficient fare differentiation. Any additional revenues from peak fare differentiation could provide a source of funds for meeting the strains and challenges under which many systems currently operate.

"Transit ridership is largely a function of the price and service characteristics of the product that is delivered." (Cervero 1990, p117)

What are the impacts, opportunities and constraints of different ticketing technologies?

Ticketing technology is a fundamental issue in addressing peak period congestion through pricing-based responses that address the "*…increased practicality of new fare collection technology (ie contactless smart card technologies) to support higher levels of fare complexity…"* (Streeting & Charles 2006). Without intelligent application of technology the effect of any pricing-based initiative will be obviated, or in some scenarios fare structure-based initiatives may simply not be possible in the first instance.

The implementation of optimal pricing structures increases the complexity of calculating fares for individual passengers based on time, distance, rail line, and station choice (as well as concession rates and other factors). Paper and magnetic stripe tickets are generally unable to meet the demands of this greater complexity of fare calculation. The persistence of paper tickets as a common payment method may indeed be a prime *cause* behind over-simplified and non-optimal fare structures (Streeting & Charles 2006).

Contactless smartcard technologies offer the ability to implement fully differentiated fare regimes that address effective pricing for rail travel (addressing distance, station choice, rail line and time of travel). This is because the technology effectively facilitates the reasonably complex fare calculations required (Streeting & Charles 2006; Li & Wong 1996).

A weak-point in the use of smartcards for fully differentiated fare structures would be in the quality of information and pricing feedback provided to customers. Current hardware provides limited pricing information to customers, in the form of a read-out that is flashed onto a small LCD screen for a short period while the customer is in motion. If differentiated fare structures are to be anything more than revenue-raising measures, they would need to provide clear pricing *signals* that elicit beneficial shifts in travel behaviour into less congested options.

A suite of pricing information measures would also need to be in place to reinforce the pricing differentials by ensuring that customers are fully aware of the travel choices they are making and the choices that are available. A range of potential initiatives include signage and advertising, information packages, online information and face-to-face contact with agency staff to deliver the message that customers can minimise their fare costs by choosing wisely in terms of travel time, route and choice of end-station. To paint an anecdotal example – in a location such as Sydney, these considerations might one day articulate themselves through clear signage and information carrying the message "*Customers please be aware – Town Hall is a full-fare station*" or alternatively, "*Museum is a discounted fare station* … " as the case may be.

> *"Overall … the 3 Ps principle … product, price and promotion – holds equally well for mass transit." (Cervero 1990, p136)*

F. SUMMARY OF CHALLENGES ASSOCIATED WITH PEAK-LOADED DEMAND PROFILES

Overloading of rail systems during peak periods presents a range of problems including: train crowding, station crowding, corridor capacity constraints, underutilised infrastructure on "return" journeys, and underutilised infrastructure in non-peak journeys and periods.

Crowding on trains

The issue of rail transit passenger discomfort on overcrowded trains is of particular concern to rail operators seeking to attract passengers and compete with other travel options by offering superior convenience and service at a given price (Vuchic 2005; TCRP100 2003). Safety concerns are another fundamental consequence (Li & Wong 1996; Doggett et al 2004). For travellers themselves, both of these aspects are magnified and personalised, while any "benefits" of heavy loadings to rail operators (in the form of solid revenue streams) are only of abstract interest to passengers at best. Overcrowding reduces the ability of rail to attract new passengers and source new revenue streams – and in this sense, any short-term revenue-growth signals that increased crowding might send to operators are in fact a distracting influence on effective medium-term rail planning.

> *"Passenger response …found standing and crowded seating to increase the cost of rail travel as perceived by passengers." (Douglas & Karpouzis 2005)*

In a cost-oriented sense, overcrowding of trains becomes a pressure point on a transit agency that may articulate itself through demand for new rolling stock, offering higher-capacity design. The potential costs associated with demand for new trains are obvious (Brooker & Moore 2008).

Station overcrowding

More or less the same set of problems associated with train crowding are present at congested railway stations, and create similar disincentives to use rail transit, while also leading to serious issues of passenger and staff safety (Vuchic 2007). As many stations are constrained by configuration and layout, the range of infrastructure-based response options are limited (station expansion is uncommon, especially in CBD locations). Heavily loaded stations can compound rail reliability problems if passenger loading becomes inefficient and time-consuming, as it invariably does under overcrowded conditions. Station-throughput is the most common “weakest link” in overall system capacity, with platform loading the most common weak link in station capacity (Vuchic 2005).

In a cost-oriented sense, overcrowding of stations becomes a pressure point on a transit agency that may articulate itself through demands for station redesign and/or expansion, or even the opening of new stations.

Corridor capacity constraints

Limitations on the ability of particular corridors to continue growing revenue by attracting additional passengers are an obvious consequence of constrained capacity and congested rail lines (Booker & Moore 2008).

> *“On most lines leading to the Sydney CBD, the high level of crowding of peak hour train services, which is now typically 100-150% seated capacity, combined with the limited time for which passengers are prepared to stand while travelling on trains, means that effective growth in peak hour rail passenger travel is now only possible from areas within “standing commuter” travel distance from the Sydney CBD.” (Douglas & Karpouzis 2005)*

Upgrading of capacity on rail lines requires either additional tracks or better train management systems – with significant cost implications.

Poor overall utilisation of infrastructure, rolling stock and system capacity

A key objective of rail operators is that their network be utilised in an efficient and cost-effective manner (Vuchic 2005; TCRP100 2003). High passenger loadings during peak periods *appear* to be “efficient” and obviously are a mainstay of ticketing revenue for many systems. Adopting a narrow point of view might tempt an observer to suggest that full trains during peak periods, carrying large numbers of paying customers and covering the costs of staff, real estate and equipment during a particular inward journey, all add up to a strong outcome for the rail operator. But this would be a shortsighted and incorrect assumption, as every inward morning journey to the CBD also involves a return trip generating little in the way of network utilisation or ticketing revenues.

The ratio or balance between “base” loadings or demand levels and “peak” demand levels is an indicator of the overall health of a rail network (Vuchic 2005). This ratio will reflect the relative effectiveness at which travel demand utilises offered services and runs. It reflects the overall loading and utilisation of expensive rolling stock, and the question of whether stations and rail corridors are being worked at close to their full potential in a sustained manner. The financial issues associated with the peak-loaded network paradigm have been summarised as problems of poor “cost per passenger” outcomes. Whereas the “peak direction cost per passenger” may look attractive, the “whole picture” is far from ideal.

Here we see the “paradox” at the heart of excessively peak-loaded rail systems: an entire rail system becomes geared toward peak capacity, yet this available capacity is *inefficiently utilised* when a holistic viewpoint is adopted, and when the demand/capacity equation for the non-peak period or direction of travel is also considered.

G. RECOMMENDATIONS FOR ADDRESSING PEAK PERIOD CONGESTION

The list of options for addressing peak period congestion includes:

- Increasing capacity during the peak through intelligent operational planning, rolling stock and infrastructure responses
- Improved off-peak service levels with the aim of shifting trips
- Differential pricing with the aim of shifting trips – potentially through increases in peak prices as well as decreases in the off-peak
- Shifting station choice away from overloaded stations
- Developing a wider set of peak destinations over time
- Communication-based measures.

It seems that exploration of the potential costs and benefits of these demand management tools is set to grow in coming years. Tools and techniques offering effective impacts without huge cost implications are likely to receive priority, but the full range of options, including capacity-expansion are not to be ruled out.

Pricing initiatives are likely to be a mainstay of demand management efforts, due to their reasonable level of effectiveness and the ability to implement them in either revenue-neutral or revenue-positive approaches. Better tracking and data on peak period and demand/capacity-related problems should be a starting point. Further research into the motivations and needs of regular rail travellers will yield a better understanding of the options available and their likely effectiveness.

Paying increased attention to the methods, tools and techniques in use in major rail systems throughout Australia and internationally will identify innovative options that could have positive impacts. Greater exchange of information between agencies on rail demand management strategy is a cost-effective means of generating change and improved outcomes. A posture of "active management" of demand will be required of rail transit agencies into the future.

BIBLIOGRAPHY

Australian Transport Council (ATC) (2006) *National Guidelines for Transport System Management in Australia – Part 4 Urban Transport* Canberra: Australian Transport Council (author and publisher)

Brooker, T and Moore, S (2008) *Recent developments in rail passenger travel demand and transit oriented development in Sydney* Australian Transport Research Forum 2008

Cervero, R (1998) *The Transit Metropolis – A Global Enquiry* Washington DC: Island Press

Cervero, R (1990) Transit pricing research *Transportation.* Edition 17 - 1990

City of Munich *Nahverkehrsplan der Landeshauptstadt München.* (translation – "Public Transport Plan for the State Capital of Munich") (July 2005) City of Munich – Department of Urban Planning and Building Regulation. Munich, Bavaria

Doggett, M Douglas, N and McGregor, G (2004) *The spatial profile of trips within the Sydney CBD* Australian Transport Research Forum 2004

Douglas, N and Karpouzis, G (2005) *Estimating the cost of train overcrowding* Australian Transport Research Forum 2005

Li, S and Wong, F (1994) The effectiveness of differential pricing on route choice *Transportation.* Edition 21 - 1994

Lyons, G and Chatterjee, K (2008) A human perspective on the daily commute: costs, benefits and trade-offs *Transport Reviews.* Volume 28, No. 2, March 2008

Streeting, M and Charles, P (2006) *Developments in transit fare policy reform* Australian Transport Research Forum 2006

Thevathasan, A, and Balachandran, B (2007) *Customer's perceptions of metropolitan train services in Melbourne* Australian Transport Research Forum 2004

TNS Social Research (TNS) (2008) *Peak to off-peak – encouraging the shift* Sydney: TNS (author and publisher – report for CityRail)

Transit Co-operative Research Program (TCRP) Report 100 (2003) *Transit Capacity and Quality of Service Manual* Washington DC: Transportation Research Board/National Academy Press

Transit Co-operative Research Program (TCRP) Report 94 (2003) *fare policies, structures and technologies: update* Washington DC: Transportation Research Board/National Academy Press

Transit Co-operative Research Program (TCRP) Report 88 (2003) *A guidebook for developing a transit performance-measurement system* Washington DC: Transportation Research Board/National Academy Press

Transit Co-operative Research Program (TCRP) Report 13 (1996) *Rail Transit Capacity* Washington DC: Transportation Research Board/National Academy Press

Vuchic, V R (2005) *Urban Transit – Operations Planning and Economics* New Jersey: John Wiley & Sons

Vuchic, V R (2007) *Urban Transit – Systems and Technology* New Jersey: John Wiley & Sons